RAW SEQUENCE LISTING
PATENT APPLICATION *US/09/003,574*

INPUT SET: S23591.raw

This Raw Listing contains the General Information Section and up to the first 5 pages.

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Tripp, Cynthia Ann
Frank, Glenn R.
Grieve, Robert B.

(ii) TITLE OF INVENTION: NOVEL PARASITE ASTACIN METALLOENDOPEPTIDASE PROTEINS

(iii) NUMBER OF SEQUENCES: 36

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SHERIDAN ROSS P.C.
(B) STREET: 1700 LINCOLN ST., SUITE 3500
(C) CITY: DENVER
(D) STATE: CO
(E) COUNTRY: USA
(F) ZIP: 80203

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Connell, Gary J.
(B) REGISTRATION NUMBER: 32,020
(C) REFERENCE/DOCKET NUMBER: 2618-21-1-C1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (303) 863-9700
(B) TELEFAX: (303) 863-0223

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1299 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTTTTTTTTT TTTTTTTTGT TTCATTGTTC AGTCAGTGGA AAATTATCGA ACGCAGAAAG 60

CATCACGAAA TACGTTAGAT CACATCAAAC AACTTATCAC CTTGAACGTA CAAAGAGAGA 120

TTGGAAACAT AGATGATAAG ACATTAGCTG ATGAAATAGT ATTACAACGA CGGGATCCTG 180

AGGCAAAATG GCATCATAAT GAACTATTCA TTAATGATCC AGATGCATAC TATCAAGGCG 240

ATGTCGATTT GTCGGAAAAA CAAGCCGAAA TTCTAAGCGA ACATTTTAAA AATGAAATTG 300

CTTTAACAGA GAAAGACGAC ACAATAATAC GGCGAAAAAA GAGCATTGGT CGTGAACCAT 360

TTTACGTAAG ATGGAATCAT AAACGTCCCA TTAGCTATGA ATTTGCGGAA AGTATTCCAT 420

TAGAAACACG TAGAAAAATT CGTTCAGCAA TAGCAATGTG GGAAGAACGA ACATGCATAC 480

GATTCCAAGA AAATGGCCCA AATGTAGATC GAATTGAATT TTACGACGGT GGCGGTTGTT 540

CAAGTTTTGT CGGCCGAACA GGAGGGAATT TCAATTTCAA CACCAGGATG TGATATTATT 600

GGTATTATAT CACATGAAAT TGGTCATACT TTAGGAATAT TTCATGAGCA AGCACGTCGT 660

GATCAAAAAA ATCATATTTT TATTAATTAC AACAATATTC CATCAAGCCG TTGGAACAAT 720

TTTTTTCCAT TATCAGAATA TGAAGCTGAT ATGTTTAATT TACCTTATGA TACAGGATCA 780

GTAATGCACT ATGGTTCATA CGGATTTGCA AGAAATCCGT ATGAACCAAC TATTACAACA 840

CGTGATAAAT TTCAACAGTA CACAATTGGG CAACGTGAAG GGCCATCATT TCTGGATTAT 900

GCATCTGTTA AGCTTTATCT ACAAACGCAT TAATGATATT GTTATCAAAT GGATGATAAT 960

TTCAATAAGT ATAAACAGCG CTTATCGTTG TACAGAACAA TGTGCTGATA TGCACTGCGA 1020

TCATAATGGT TATCCGGATC CTAATAATTG CGCGAAATGC TTGTGTCCAG ATGGTTTTGC 1080

TGGTCGTACC TGTCAATTTG TTCAATATAC ATCTTGCGGA GCTCTCATTA AGGTAAGTAT 1140

TGTCTTTTGA CCTCTTCTCT GACTAAAATA TAAGTTAAGC ATATGTATCT TCCGTCTAAT 1200

GATTTTCTTG ATTTTGATTT GTTCAATGCT CTTCTTGATA ATAATATAAA AATTTTTGAA 1260

AATAAAGTTA ACTTTTGGTC AAAAAAAAAA AAAAAAAAA 1299

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAAAGCATCA CGAAATACGT TAGATCACAT CAAACAACTT ATCACCTTGA ACGTACAAAG 60

AGAGATTGGA AACATAGATG ATAAGACATT AGCTGATGAA ATAGTATTAC AACGACGGGA 120

TCCTGAGGCA AAATGGCATC ATAATGAACT ATTCATTAAT GATCCAGATG CATACTATCA 180

AGGCGATGTC GATTTGTCGG AAAAACAAGC CGAAATTCTA AGCGAACATT TTAAAAATGA 240

AATTGCTTTA ACAGAGAAAG ACGACACAAT AATACGGCGA AAAAAGAGCA TTGGTCGTGA 300

ACCATTTTAC GTAAGATGGA ATCATAAACG TCCCATTAGC TATGAATTTG CGGAAAGTAT 360

TCCATTAGAA ACACGTAGAA AAATTCGTTC AGCAATAGCA ATGTGGGAAG AACGAACATG 420

CATACGATTC CAAGAAAATG GCCCAAATGT AGATCGAATT GAATTTTACG ACGGTGGCGG 480

TTGTTCAAGT TTTGTCGGCC GAACAGGAGG GAATTTCAAT TTCAACACCA GGATGTGATA 540

TTATTGGTAT TATATCACAT GAAATTGGTC ATACTTTAGG AATATTTCAT GAGCAAGCAC 600

GTCGTGATCA AAAAAATCAT ATTTTTATTA ATTACAACAA TATTCCATCA AGCCGTTGGA 660

ACAATTTTTT TCCATTATCA GAATATGAAG CTGATATGTT TAATTTACCT TATGATACAG 720

GATCAGTAAT GCACTATGGT TCATACGGAT TTGCAAGAAA TCCGTATGAA CCAACTATTA 780

CAACACGTGA TAAATTTCAA CAGTACACAA TTGGGCAACG TGAAGGGCCA TCATTTCTGG 840

ATTATGCATC TGATAAACAG CGCTTATCGT TGTACAGAAC AATGTGCTGA TATGCACTGC 900

GATCATAATG GTTATCCGGA TCCTAATAAT TGCGCGAAAT GCTTGTGTCC AGATGGTTTT 960

GCTGGTCGTA CCTGTCAATT TGTTCAATAT ACATCTTGCG GAGCTCTCAT TAAGGCGAGG 1020

AAAATGCCTG TTACGATTTC GAGCCCAAAT TATCCAAACT TCTTCAATGT TGGTGATCAA 1080

TGTATTTGGT TGCTTACAGC TCCACGCGTG ATTCGTAAAT TTGCAGTTTG TTGAACAATT 1140

153 TCAATTACAA TGTGAAGATA CGTGTGATAA ATCCTATGTA GAAGTGAAAG CTGACGCTGA 1200
154
155 TTTTCGACCT ACTGGATATC GATTTTGTTG TTCGCGAGTG CCACGTCATA TTTTTCAATC 1260
156
157 TGCGACAAAC GAGATGGTAG TAATATTTCG CGGTTTTGGT GATGCGGGAA ATGGCTTTAA 1320
158
159 AGCTAAAATT TGGTCAAACG TAGATGATGA TATAGCTAAT ACAATTGTAA CAACTGAAAT 1380
160
161 GGCAAAAATT TCGGAAAAAA TACCGAAGCT AACAGTTCCA ATAGTTAAAA CTATTACCAC 1440
162
163 TCCTACAATA ACAACTACTA CTGCTTTCAT GATATCACCC AAGAAAGGCA ATGTCACCGC 1500
164
165 CACGAGAGTT GCTATCACTA CTACGCCGAC TACTACAATT ACTACGACTA TTGCCGGTAC 1560
166
167 GTACCAATCA CCGTAACTAA TAATACTACA CCTGTAGTAA GTGAAACTTT ACCATCATTG 1620
168
169 CCAGTCAAGA TTCGAAACAA AATAGGTGCA TGCGAATGTG GTGAATGGAC AGAATGGACA 1680
170
171 GGTCCATGCT CTCAAGAATG TGGCGGTTGC GGAAAACGTC TTCGAACACG TCAGTGTTCA 1740
172
173 TCAGATACGG AATGTAGAAC AGAAGAAAAA CGTGCGTGTG CTTTTAAGTT TGCCCATACG 1800
174
175 GGACTAATTT CCTTATCAAT AATGGAGAGT TTCATATACT TTGGAAGGGC TGCTGTGTTG 1860
176
177 GTCTATTCCG ATCGGGAGAT ATGTGTTCAG CACTTGATGA TAACGAGAAT CCATTTCTGA 1920
178
179 AATTTCTAGA ATCACTGTTG AACATGCAAG ATTCTCGAAA AAACGATAAT TTGCCTGACT 1980
180
181 CGAAAAAGAA GTGATTGAAT GATTCGATAA TATTGATTAA TAAAACGGGT TGTATTCTCG 2040
182
183 TCATAGAGTA TCCGTTGATG TTTTTATCCA AAAAATTCTC TTGCTTTTAA TTATTGTGAA 2100
184
185 TAAAACTTTT GTTTACCCAA AAAAAA 2126
186
187 (2) INFORMATION FOR SEQ ID NO:3:
188
189 (i) SEQUENCE CHARACTERISTICS:
190 (A) LENGTH: 191 amino acids
191 (B) TYPE: amino acid
192 (C) STRANDEDNESS:
193 (D) TOPOLOGY: linear
194
195 (ii) MOLECULE TYPE: protein
196
197
198
199 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
200
201 Cys Phe Ile Val Gln Ser Val Glu Asn Tyr Arg Thr Gln Lys Ala Ser
202 1 5 10 15
203
204 Arg Asn Thr Leu Asp His Ile Lys Gln Leu Ile Thr Leu Asn Val Gln
205 20 25 30

Arg Glu Ile Gly Asn Ile Asp Asp Lys Thr Leu Ala Asp Glu Ile Val
35 40 45

Leu Gln Arg Arg Asp Pro Glu Ala Lys Trp His His Asn Glu Leu Phe
50 55 60

Ile Asn Asp Pro Asp Ala Tyr Tyr Gln Gly Asp Val Asp Leu Ser Glu
65 70 75 80

Lys Gln Ala Glu Ile Leu Ser Glu His Phe Lys Asn Glu Ile Ala Leu
85 90 95

Thr Glu Lys Asp Asp Thr Ile Ile Arg Arg Lys Lys Ser Ile Gly Arg
100 105 110

Glu Pro Phe Tyr Val Arg Trp Asn His Lys Arg Pro Ile Ser Tyr Glu
115 120 125

Phe Ala Glu Ser Ile Pro Leu Glu Thr Arg Arg Lys Ile Arg Ser Ala
130 135 140

Ile Ala Met Trp Glu Glu Arg Thr Cys Ile Arg Phe Gln Glu Asn Gly
145 150 155 160

Pro Asn Val Asp Arg Ile Glu Phe Tyr Asp Gly Gly Gly Cys Ser Ser
165 170 175

Phe Val Gly Arg Thr Gly Gly Asn Phe Asn Phe Asn Thr Arg Met
180 185 190

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ile Glu Leu Asn Phe Thr Thr Val Ala Val Val Gln Val Leu Ser Ala
1 5 10 15

Glu Gln Glu Gly Ile Ser Ile Ser Thr Pro Gly Cys Asp Ile Ile Gly
20 25 30

Ile Ile Ser His Glu Ile Gly His Thr Leu Gly Ile Phe His Glu Gln
35 40 45

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/09/003,574*

DATE: 02/20/98
TIME: 09:40:56

INPUT SET: S23591.raw

Line	Error	Original Text